Stats 101B: 2^3 Factorial Research on Blood Glucose

Xinyang Wang, Janel Wu, Yichen Wang,
Jialing Yuan, Yuexuan Wu, Mindy Zhu

June, 2024

1 Introduction

The increasing prevalence of metabolic disorders worldwide prompts an urgent need for tailored lifestyle interventions. Our study is motivated by the need to understand how dietary intake, physical activity, and caffeine consumption independently and collectively influence blood glucose levels. Given the varying responses across different age groups, integrating age as a blocking factor allows for a more nuanced analysis of these lifestyle factors. This research aims to unravel the complex interactions between food, exercise, and caffeine to inform more personalized health recommendations that could effectively manage or mitigate glucose fluctuations.

This project expects to highlight significant lifestyle factors that affect metabolic health, particularly blood glucose levels, across different age demographics. By conducting a controlled three-factorial study, we aim to isolate the effects of each factor and their interactions. The results are anticipated to contribute substantially to the existing literature on metabolic health, providing a foundation for developing age and lifestyle-specific dietary and exercise recommendations.

2 Design of the experiment

In order to explore the factors that affect the response variable, namely the blood glucose levels of islanders, we conducted a 2^3 factorial design with two blocks. We selected three factors: coffee, food, and exercise, and each factor has two levels: coffee (regular, decaf), food (70% dark chocolate, 50g of white chocolate), and exercise (5-min light jogging, 1-km running). We then put observations with age smaller than 55 into block 1 and observations with age larger than or equal to 55 into block 2, which makes age a blocking factor. With 32 observations in total, we set the first 16 observations as block 1 and the remaining 16 observations as block 2, with each block containing two replicates.

Now we introduce the detailed procedures. We collected samples using cluster sampling that treats each island as a cluster, and within each island, further stratify by areas. We have 16 groups and 2 replicates each, thus in total 32 observations. We allocated 11 samples to Island Ironbard and Island Bonne Santé and 10 samples to Providence. Within each island, participants are randomly assigned to a city by numbering the cities and using a random number generator. Then, as the houses in the cities were already numbered on the Islands, a random number generator was used to select the houses of individuals asked for consent. Only citizens over the age of 18 were used, and if a person declined, the random number generator was used again to repeat the process. Since we have age as a block, such randomization will contain 2 groups, with 16 people under 55 and 16 under 55. If randomization of either group is satisfied, then the remaining assignment will focus on the other group. This randomized sampling was used to most accurately represent the entire population of

The Islands and eliminate potential biases as well as the effects of confounding variables. However, it is important to note that the samples were not completely random as either a male or female was intentionally selected from their household to attain roughly equal numbers of both genders in the overall sample. After randomly selecting each individual, we first tested their glucose levels, experimented with different activities, and collected their glucose levels in the end. Different activities include drinking 250 ml of coffee or 250 ml of decaffeinated coffee, eating 50g of 70% dark chocolate or 50g of white chocolate. Per Island's rule, we know that 1 year on the island is equivalent to 28 days in the Gregorian calendar. Since the blood glucose level usually drops after 3 hours, which is equivalent to 13 minutes in Gregorian calendar, we let all samples to rest for 13 minutes for the glucose levels to stabilize after ingesting food. In the end, they did light jogging for 5 minutes or running 1 km.

After the experiment, we want to make sure that we have used enough number of replications to make the test have more power. We did a power test with ANOVA using $k = 16$ and $n = 2$. The result is a power of 1, implying that we can conclude the treatment means are significantly different when the true means are different. We can also tell that food factor is statistically significant. Our sample size, which equals 32, is appropriate to test this experiment.

renewed design(final blood glucose)							
			Block1:age<55		block2: age >=55		
coffee	food	exercise	replicate1	replicate2	replicate1	replicate2	mean
-	-	-	116	116	106	110	112
-	+	-	129	131	114	120	123.5
+	-	-	111	104	115	104	108.5
+	+	-	136	114	120	118	122
-	-	+	106	113	110	107	109
-	+	+	114	132	122	116	121
+	-	+	101	111	104	105	105.25
+	+	+	112	138	131	110	122.75

Figure 1: Data collection

3 Results and Interpretation

```
                    Df Sum Sq Mean Sq F value   Pr(>F)    
coffee               1   24.5    24.5   0.395    0.535    
food                 1 1485.1  1485.1  23.962 5.42e-05 ***
Exercise             1   32.0    32.0   0.516    0.479    
coffee:food          1   28.1    28.1   0.454    0.507    
coffee:Exercise      1    4.5     4.5   0.073    0.790    
food:Exercise        1   10.1    10.1   0.163    0.690    
coffee:food:Exercise 1    6.1     6.1   0.099    0.756    
Residuals           24 1487.5    62.0                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 2: Summary of ANOVA table

Note: All plots not shown are in the appendix. We first analyzed the data as a whole without blocking(Figure 2). From summary of ANOVA table above, we know there is only one significant factor: food. It contributes mostly with high mean square value and very low p-value, indicating how largely food can affect blood glucose level. However, other factors and interactions are all non-significant by comparison.

Figure 6 shows corresponding magnitude and direction of estimated factor effects. While food has a large positive value, coffee and exercise actually decrease blood glucose level on a slight scale.

```
Coefficients:
                     Estimate
(Intercept)          115.5000
coffee                -0.8750
food                   6.8125
Exercise              -1.0000
coffee:food            0.9375
coffee:Exercise        0.3750
food:Exercise          0.5625
coffee:food:Exercise   0.4375
---
```

Figure 3: Regression coefficients

The coefficients in regression model (Figure 3) are shown above, which are half of estimated factor effects. When only considering significant effects, the model is $y = 115.50 + 6.8125*x$, where x is the coded variable for food and y is final blood glucose level. This indicates an average of 6.8125 change in blood glucose level for changing food type from low (dark chocolate) to high (white chocolate), which follows intuition that white chocolate is 'unhealthier' and has higher sugar level. However, since there are only three factors in experiment, this model and interpretation may not be appropriate or practical to apply a in real world as glucose level can be affected by other factors that we did not discuss.

Stepping into the model assumption, we first check the residuals plot (Figure 7). There exists a slightly discernible pattern that as fitted (predicted) values increase, corresponding residuals becoming diverge. Overall, the pattern is symmetric, but further study can be done to collect more data and determine a more reliable pattern.

When we check normality plot(Figure 8), however, it shows a straight slope and few outliers, indicating the model normality assumption is well satisfied.

Though the experiment includes two replicates and normal probability plot is not necessary, we still conduct one to reassure our former analysis (Figure 9). From the plot, we recognize that food is significantly far away from the line, which again indicates its importance.

Figure 10 shows main effect plots of three factors. We observe that while slope is the same, y axis has different scale. As food changes level, the corresponding y axis changes from less than 110 to over 120, demonstrating its large positive magnitude. The other two factors follow the data of estimated effects and appear to be less influential on y.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
coffee	1	24.5	24.5	0.425	0.521	
food	1	1485.1	1485.1	25.770	3.86e-05	***
Exercise	1	32.0	32.0	0.555	0.464	
block	1	162.0	162.0	2.811	0.107	
coffee:food	1	28.1	28.1	0.488	0.492	
coffee:Exercise	1	4.5	4.5	0.078	0.782	
food:Exercise	1	10.1	10.1	0.176	0.679	
coffee:food:Exercise	1	6.1	6.1	0.106	0.747	
Residuals	23	1325.5	57.6			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figure 4: Summary of ANOVA–with blocking

We then conduct ANOVA analysis with age as blocking(Figure 4). Even though the p-value is 0.107 which is not significant, it's near 0.05 and has second highest mean square, so it does have some contribution on blood glucose level, no matter negative or positive. This shows that different ages are somewhat influential when measuring blood glucose level in experiment. This may be further investigated by enlarging sample size or categorizing ages in more different age groups such as 10-20, 20-30, etc. so as to confirm detailed impact of age factor.

From the interaction plot (Figure 11), we see some small interactions between three factors. From interactions of AB, AC, and BC, we learn that in order to reach the minimized blood glucose level, the best combination would be A high(coffee), B low (dark chocolate), and C high (running for 1 km). We can see the intuition that dark chocolate and more intense exercise have a somewhat smaller increase in blood glucose level.

As for power calculation(Figure 12), we got a power of 1 due to a large effect size, which shows a really significant difference of treatment means from null hypothesis. As the above analysis all depend on final glucose level, they contain a potential problem of ignoring each person's baseline blood glucose level. It's possible that some people have diabetes and own a higher initial glucose level. In this case, only final glucose level is not reliable. Therefore, we also did analysis on change of blood glucose and found pretty similar results. The following are the R results from this analysis(Figure 5, Figure 13-17)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
coffee	1	1.5	1.5	0.028	0.869	
food	1	1212.8	1212.8	22.155	8.75e-05	***
Exercise	1	7.0	7.0	0.128	0.723	
coffee:food	1	52.5	52.5	0.960	0.337	
coffee:Exercise	1	13.8	13.8	0.252	0.620	
food:Exercise	1	7.0	7.0	0.128	0.723	
coffee:food:Exercise	1	30.0	30.0	0.549	0.466	
Residuals	24	1313.7	54.7			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figure 5: ANOVA:change

4 Discussion

4.1 Summary

To summarize, our study aimed to investigate the influence of various factors on islanders' glucose levels, particularly focusing on dietary intake (food), caffeine consumption (coffee), and physical activities (exercise). Through a comprehensive analysis using ANOVA tables, both with and without blocking, it was determined that food is the most and the only statistically significant factor contributing to the influence of glucose levels. The main effect plots revealed that food consumption increases glucose levels, while coffee consumption and exercise decrease them. Specifically, from the interaction plots, we can further conclude that eating dark chocolate, drinking regular coffee, and engaging in more intensive exercise minimize glucose levels, which could be used as recommendations for islanders with diabetes. On the other hand, in terms of blocking, we used age as the block variable in our experiment. However, since the block effect is not significant, indicating it is relatively small, we can conclude that blocking is not necessary. This suggests that age does not have a substantial association with glucose levels. The reliability of our model was confirmed through the normal Q-Q plot, supporting the validity of our findings and conclusions.

4.2 Real-World Implications

Research from others in the real world also further corroborates with the results we concluded from our experimental study. The consumption of food and physical activity are some of the most studied factors relating to its impact on blood glucose (Doherty and Greaves, 2015). The results from our research added further evidence showcasing the significance that food has on a person's blood sugar levels, while also excluding other possible factors like caffeine being a significant factor as well.

Our experiment used different types of chocolate as the actual factor representing food consumption and found how dark chocolate increases blood glucose levels to a lesser degree compared with white chocolate. There are studies that have found that dark chocolate induces a beneficial short-term effect on insulin sensitivity, which is related to maintaining healthy blood glucose levels and has a positive impact on cardiovascular diseases (Grassi et al., 2008; Zhang et al., 2018). White chocolate increases blood glucose levels more due to its higher sugar content, but lacks the key compounds which come from dark chocolate that also give additional health benefits (Dehghani et al., 2024). Ultimately research on blood glucose levels is important due to its impact on metabolic health and prevents metabolic disorders like diabetes from emerging.

4.3 Limitations

Based on the discussion above, our results and findings are coherent and logical. However, there exist limitations from our study that should be addressed in future research.

Firstly, one of the uncontrollable factors in the experiment is that our subjects took the same or similar tests before taking ours, which may have impacted the glucose level we measured. For example, repeated exposure to glucose intake from previous tests might have caused adaptive physiological responses, such as increased insulin sensitivity, potentially affecting the measurements we obtained. The familiarity with the testing procedure might also lead to physiological responses that were not accounted for in our experimental design.

Secondly, individual differences in the physical constitution, such as hereditary conditions like type 2 diabetes predisposition or metabolic syndrome can significantly affect glucose levels, which may impact the conclusion we gained. These genetic and biological variations among participants might confound the effects of the treatment factors we investigated. For instance, participants with a family history of diabetes might exhibit different glucose metabolism compared to those without such a history, regardless of the interventions applied in the study, thereby influencing the conclusions we drew. While our study attempted to control for various external factors, these inherent physiological differences introduce variability that needs to be considered when interpreting our findings.

5 References

[1] Dehghani, P., Taheri, F., & Asgary, S. (2024). Decoding the delights: Unraveling the health benefits of dark chocolate in comparison to white chocolate. *Functional Food Science - Online ISSN: 2767-3146*, *4*(4), 119–133. https://doi.org/10.31989/ffs.v4i4.1300

[2] Doherty, S. T., & Greaves, S. P. (2015). Time-series analysis of continuously monitored blood glucose: The impacts of geographic and daily lifestyle factors. *Journal of Diabetes Research*, *2015*, 1–6. https://doi.org/10.1155/2015/804341

[3] Grassi, D., Desideri, G., Necozione, S., Lippi, C., Casale, R., Properzi, G., Blumberg, J. B., & Ferri, C. (2008). Blood pressure is reduced and insulin sensitivity increased in glucose-intolerant, hypertensive subjects after 15 days of consuming high-polyphenol dark chocolate13. *The Journal of Nutrition*, *138*(9), 1671–1676. https://doi.org/10.1093/jn/138.9.1671

[4] Zhang, C.-X., Long, W.-Q., Ye, Y.-B., Lu, M.-S., Zhang, N.-Q., Xu, M., Huang, J., & Su, Y.-X. (2018). Effects of chocolate-based products intake on blood glucose, insulin and ghrelin levels and on satiety in young people: A cross-over experimental study. *International Journal of Food Sciences and Nutrition*, *69*(7), 882–891. https://doi.org/10.1080/09637486.2018.1426737

Appendix

```
          coffee               food           Exercise        coffee:food
          -1.750             13.625             -2.000              1.875
 coffee:Exercise      food:Exercise coffee:food:Exercise
           0.750              1.125              0.875
```

Figure 6: Estimated factor effects

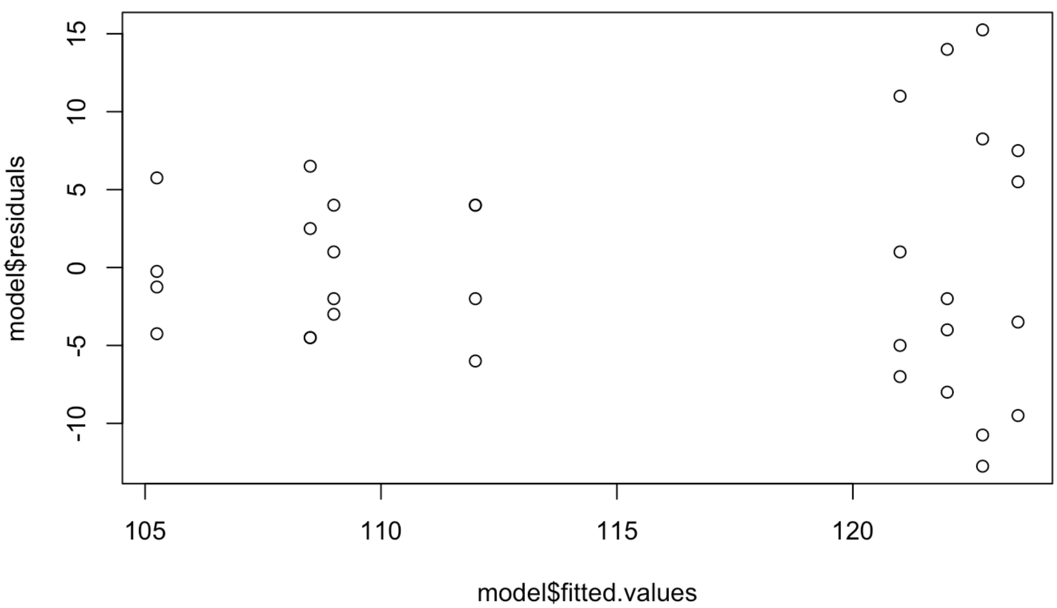

Figure 7: Residual plot

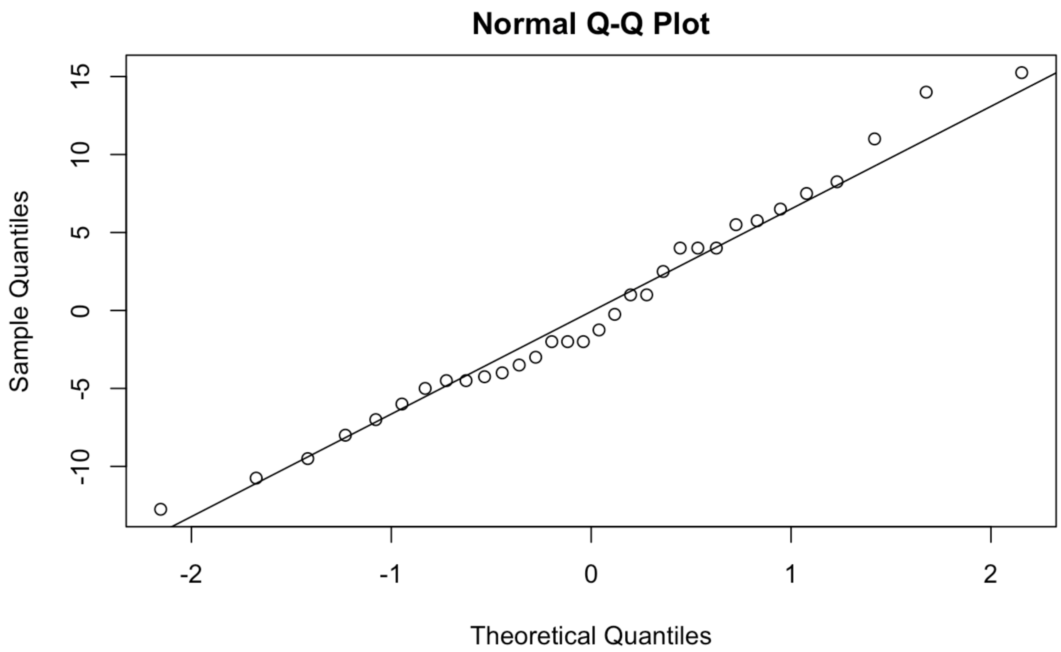

Figure 8: Normality plot

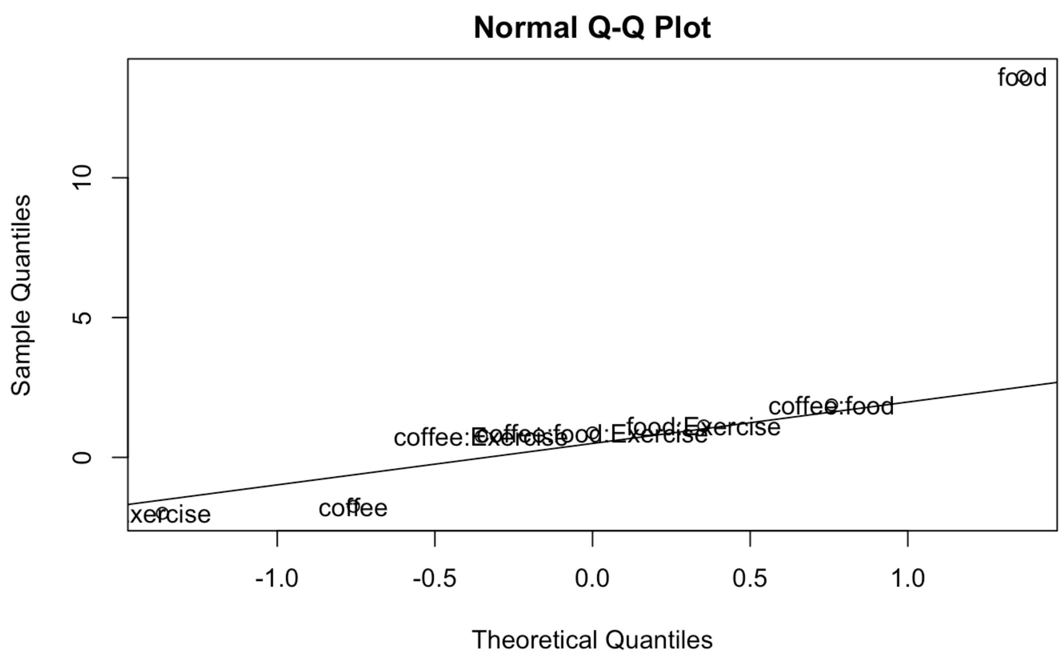

Figure 9: Normal Probability plot

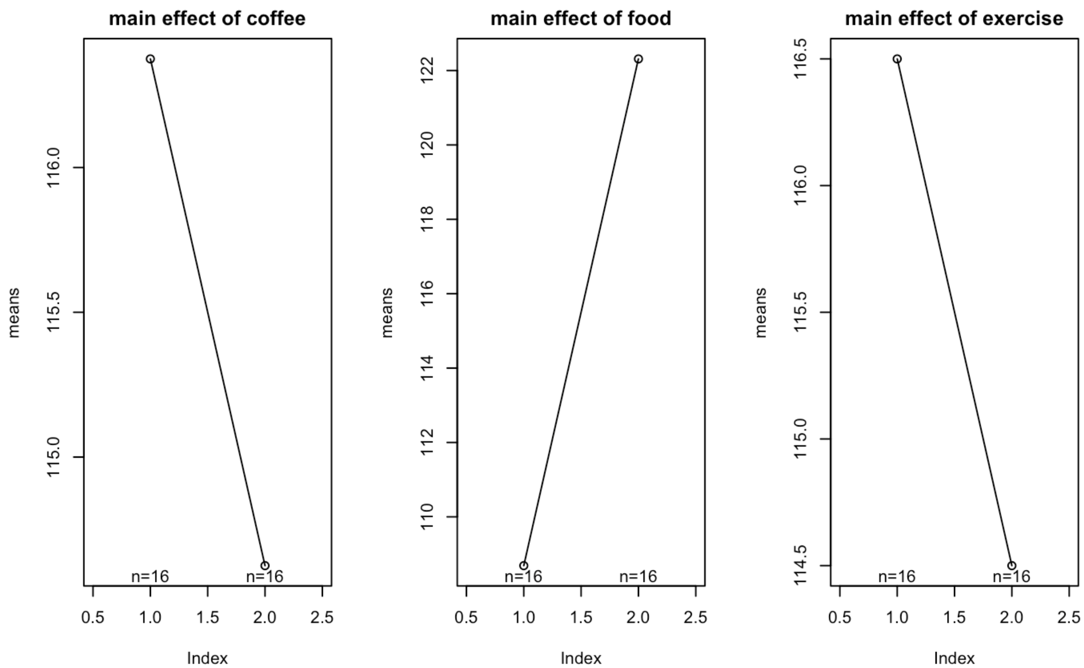

Figure 10: Main effect

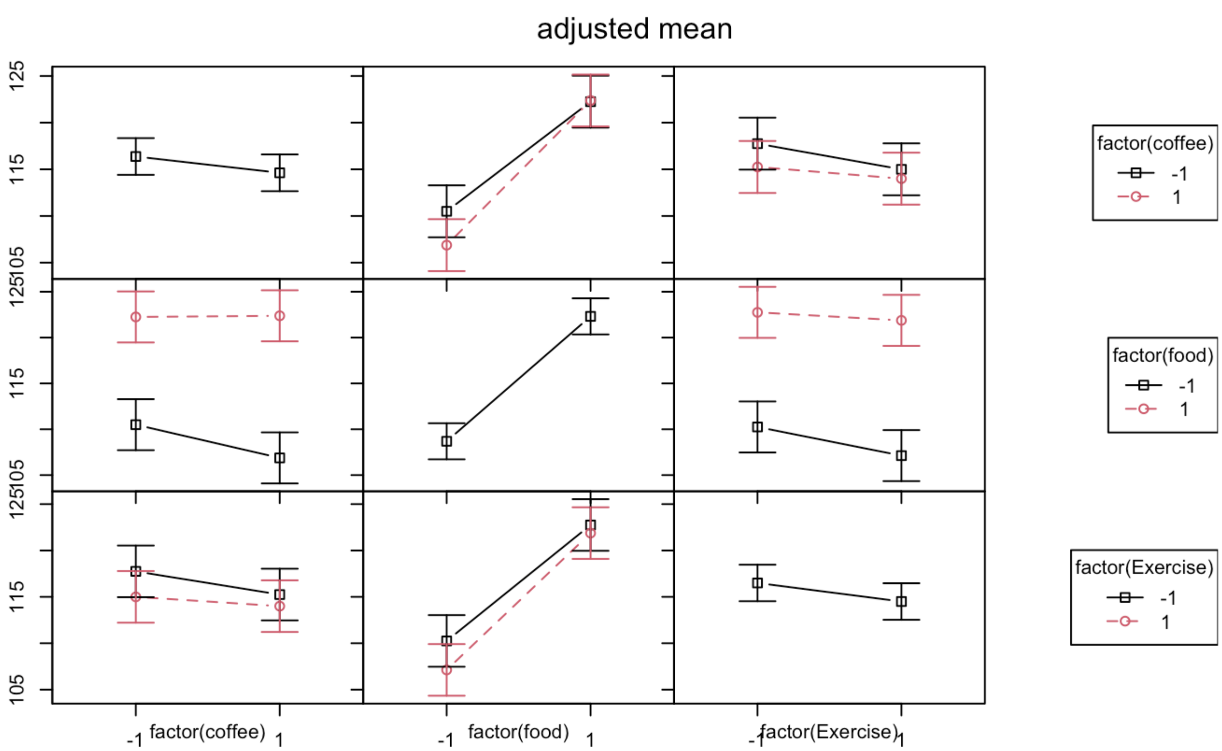

Figure 11: Interaction plot

```
     Balanced one-way analysis of variance power calculation

              k = 16
              n = 2
              f = 3.24
      sig.level = 0.05
          power = 1

NOTE: n is number in each group
```

Figure 12: Power calculation

```
             coffee                  food              Exercise           coffee:food
             0.4375               12.3125               -0.9375                2.5625
    coffee:Exercise         food:Exercise coffee:food:Exercise
             1.3125                0.9375                1.9375
```

Figure 13: Estimated effects:change

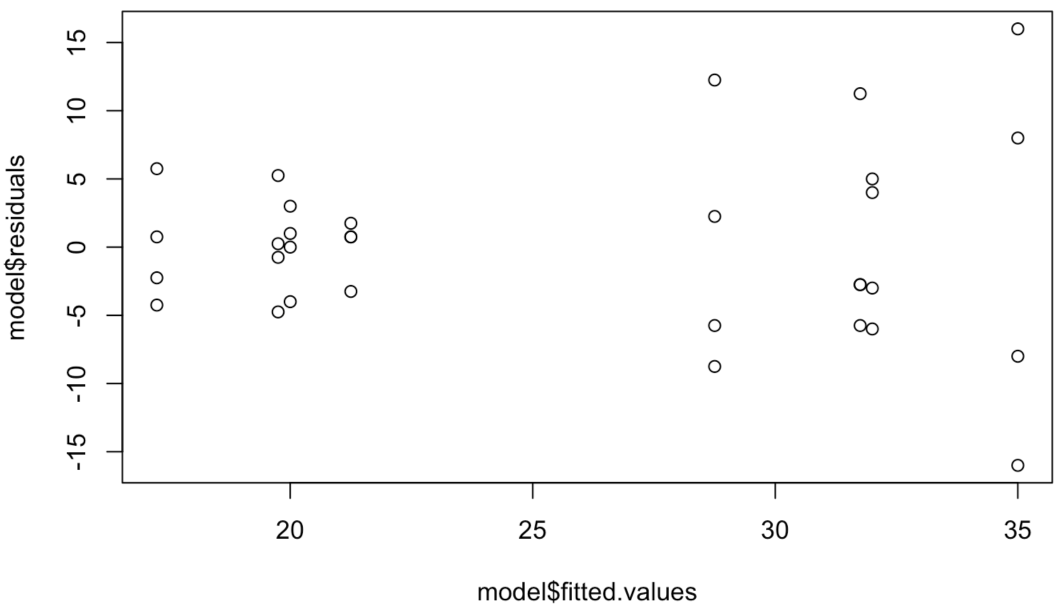

Figure 14: Residuals plot: change

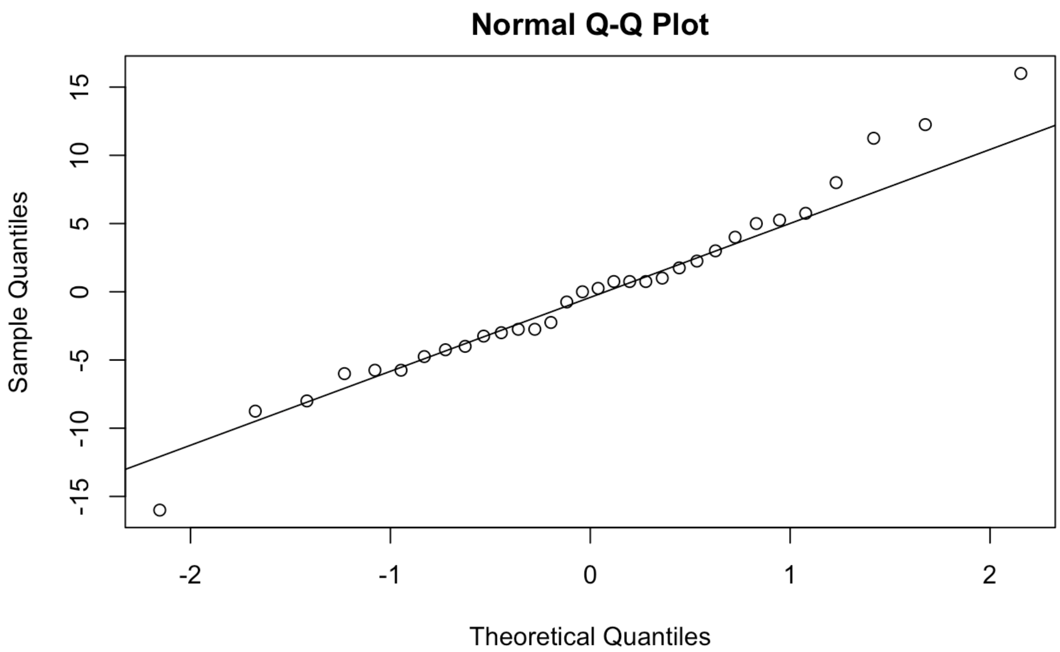

Figure 15: Normality plot:change

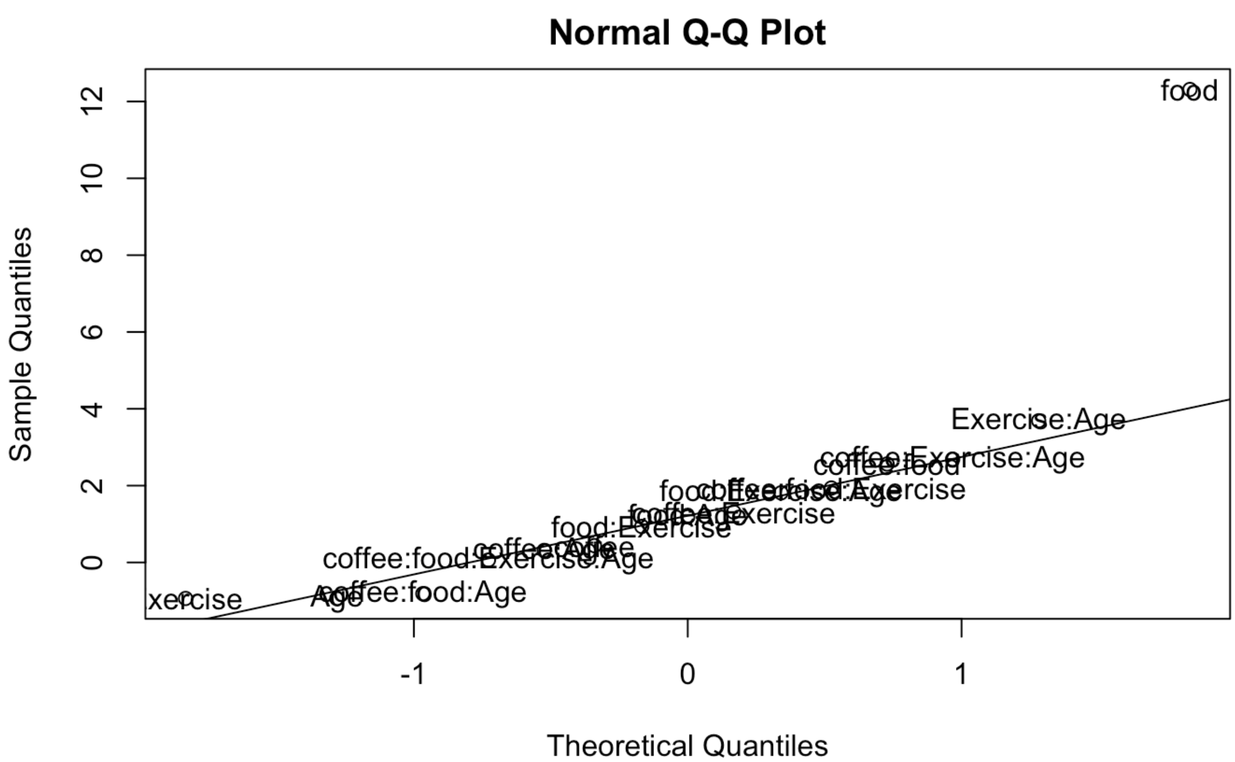

Figure 16: Normal probability: change

```
                    Df Sum Sq Mean Sq F value  Pr(>F)    
coffee               1    1.5     1.5   0.030   0.865    
food                 1 1212.8  1212.8  23.446 6.9e-05 ***
Exercise             1    7.0     7.0   0.136   0.716    
block                1  124.0   124.0   2.398   0.135    
coffee:food          1   52.5    52.5   1.016   0.324    
coffee:Exercise      1   13.8    13.8   0.266   0.611    
food:Exercise        1    7.0     7.0   0.136   0.716    
coffee:food:Exercise  1   30.0    30.0   0.581   0.454    
Residuals           23 1189.7    51.7                    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 17: ANOVA with block: change